УДК 004.652/.942

Г. Н. ЖОЛТКЕВИЧ, д-р техн. наук, Харьковский национальный университет имени В. Н. Каразина,
К. А. ФЕДОРЧЕНКО, Херсонский государственный университет

ОБ ОДНОМ КЛАССЕ КОНЦЕПТУАЛЬНЫХ МОДЕЛЕЙ

В класі напівсхем, що використовуються для концептуального моделювання інформаційних систем, знайдено підклас нерекурсивних напівсхем. Доведено, що будь-яка нерекурсивна напівсхема є схемою. Продемонстровано можливість явної побудови схеми бази даних за нерекурсивною напівсхемою, що дозволяє застосовувати запропоновану модель для автоматизації побудови та супроводження метаданих бази даних інформаційної системи.

Традиционная парадигма построения информационных систем базируется на ряде предположений, которые исключают возможность адаптации системы к структурным изменениям знаний о предметной области [1 – 2]. Статичность концептуальной модели предметной области является одним из таких предположений. Однако для широкого круга приложений такая статичность не характерна. Например, базы данных в исследовательских системах должны обладать механизмом модификации метаданных [3]. С возрастающей изменчивостью внешнего окружения организаций нестабильность метаданных становится характерной и для автоматизированных информационных управленческих систем (АИУС). В связи с этим в работе [4] была предложена алгебраическая метамодель, предназначенная для формальной спецификации концептуальных моделей предметных областей информационных систем, а в работе [5] показано, как эта метамодель может быть представлена средствами стандартных реляционных баз данных (РБД).

Как показывает анализ предметных областей управленческих информационных систем предложенный в цитируемых выше работах [4 – 5] подход к построению их моделей оказывается чрезмерно общим, не учитывающим специфику рассматриваемого класса информационных систем. Это позволяет ожидать существенного повышения эффективности обработки метаданных за счет учет такой специфики.

В связи с этим настоящая работа посвящена исследованию соответствующих сужений класса моделей, специфицируемых при помощи общей метамодели, с учетом указанной выше специфики. Подкласс моделей, который выделяется при таком сужении, назван далее авторами классом «нерекурсивных полусхем». Именно отсутствие рекурсивных определений характерно для предметных областей АИУС.

Целью настоящей работы, таким образом, является введение и исследование класса «нерекурсивных полусхем», как подкласса общего класса моделей – класса полусхем, введенного в [4].

Введем обозначения, используемые ниже.

$M_{+}(R,N)$ – множество частичных, хотя бы в одной точке определенных отображений из множества;

xR – подмножество $\{y \in Y \mid (x,y) \in R\}$ для бинарного отношения R между множествами X и Y;

X^{*} – множество конечных последовательностей элементов множества X, включающее пустую последовательность, которая всегда обозначается символом e.

$h(x_1 \dots x_k) = x_1$ – для любой непустой последовательности элементов множества X.

$t(x_1 x_2 \dots x_k) = x_2 \dots x_k$ –для любой непустой последовательности элементов множества X.

Как и в работе [5] введем понятие полусхемы предметной области следующим определением.

<u>Определение 1.</u> Полусхемой предметной области назовем тройку $S=(N,R,D)$, где N, R – конечные множества, называемые множествами понятий и ролей соответственно, $D \subset N \times M_{+}(R,N)$, для которой выполняется условие: для $n \in N$, $f,g \in nD$, $r \in \mathrm{dom}(f) \cap \mathrm{dom}(g)$ верно $f(r)=g(r)$.

Для фиксированной полусхемы $S=(N,R,D)$ как и в [4] вводятся множества именующих нитей понятий.

<u>Определение 2.</u> Элемент (n,w) множества $N \times R^{*}$ называется именующей нитью понятия n, если выполнено одно из следующих двух условий:

1. $w=e$;
2. $w = r_1 r_2 \dots r_k$ и найдется последовательность пар $(n_i, f_i) \in D$, где $i=1,\dots,k$, причем

   2.1. $n_1 = n$,

   2.2. $r_i \in \mathrm{dom}(f_i), f_i(r_i) = n_{i+1}, i=1,\dots,k-1$,

   2.3. $r_k \in \mathrm{dom}(f_k)$.

При этом в случае 1 говорят, что длина именующей нити равна 0 $(|(n,e)|=0)$, а в случае 2 – длина равна k $(|(n, r_1 r_2 \dots r_k)| = k)$.

<u>Утверждение 1.</u> Если $(n, r_1 r_2 \dots r_k)$ – именующая нить понятия n, $(n_i, f_i) \in D$, где $i=1,\dots,k$ – последовательность из определения 2,

$(n_i', f_i') \in D$, где $i = 1, \ldots, k$ – другая последовательность, для которой $n_1' = n$ и $f_i'(r_i) = n_{i+1}', i = 1, \ldots, k-1$, тогда $n_i = n_i', i = 1, \ldots, k+1$, где $n_{k+1} = f_k(r_k)$, $n_{k+1}' = f_k'(r_k)$.

Доказательство немедленно следует из определения полусхемы.

Определение 3. Если $(n, r_1 r_2 \ldots r_k)$ – именующая нить понятия n, $n' \in N$, $(n_i, f_i) \in D, i = 1, \ldots, k$ – последовательность из определения 2, причем $n' = n_i$ для некоторого $1 \le i \le k$ или $n' = f_k(r_k)$, тогда будем говорить, что именующая нить $(n, r_1 r_2 \ldots r_k)$ проходит через n'.

Из определения 3 и утверждения 1 немедленно следует

Утверждение 2. Если $(n, r_1 r_2 \ldots r_k)$ – именующая нить понятия n, $n' \in N$, $(n_i, f_i) \in D, i = 1, \ldots, k$ – последовательность из определения 2, тогда последовательность $n_1, \ldots, n_k, n_{k+1}$, где $n_{k+1} = f_k(r_k)$, определяется по $(n, r_1 r_2 \ldots r_k)$ однозначно.

Следствие. Частичное отображение $\tau : N \times R^* \to N$ корректно определено на именующих нитях понятий формулами

$$\begin{cases} \tau(n, e) = n \\ \tau(n, r_1 r_2 \ldots r_k) = n_{k+1} \end{cases},$$

где n_{k+1} – последний элемент последовательности прохождений для именующей нити $(n, r_1 r_2 \ldots r_k)$.

Определение 4. Последовательность из утверждения 2 будем называть последовательностью прохождения именующей нити.

Определение 5. Понятие $n \in N$ полусхемы $S = (N, R, D)$ называется рекурсивным, если для него существует именующая нить $(n, r_1 r_2 \ldots r_k)$, у которой в последовательности прохождения $n_1, \ldots, n_k, n_{k+1}$ найдутся $n_i = n_j$, где $1 \le i < j \le k+1$.

Определение 6. Понятие $n \in N$ полусхемы $S = (N, R, D)$ называется нерекурсивным, если оно не является рекурсивным.

Теорема 1. Понятие $n \in N$ полусхемы $S = (N, R, D)$ является нерекурсивным тогда и только тогда, когда все соответствующие ему именующие нити по длине не превосходят некоторого фиксированного натурального числа.

Доказательство необходимости.

Пусть понятие $n \in N$ полусхемы $S = (N, R, D)$ является не рекурсивным, а $(n, r_1 r_2 \ldots r_k)$ – именующая нить этого понятия. Рассмотрим для этой нити

$n_1,\dots,n_k,n_{k+1}$ – последовательность прохождения. В силу того, что понятие не рекурсивно, в последовательности $n_1,\dots,n_k,n_{k+1}$ нет повторяющихся элементов, а значит, длина ее не может превосходить мощности множества N, которую и можно выбрать в качестве ограничивающей константы.

Доказательство достаточности.

Пусть длины всех именующих нитей понятия $n \in N$ полусхемы $S=(N,R,D)$ равномерно ограниченны некоторой константой L. Предположим, однако, что существует именующая нить $(n,r_1r_2\dots r_k)$ этого понятия, в последовательности прохождения $n_1,\dots,n_k,n_{k+1}$ которой есть два совпадающих элемента, например, $n_{i_1}=n_{i_2}=n'$, $1\le i_1<i_2\le k+1$. Но тогда для произвольного $m>0$ конструкции вида

$$\left(n,r_1\dots r_{i_1-1}\left(r_{i_1}\dots r_{i_2-1}\right)^m r_{i_2}\dots r_k\right)$$

будут именующими нитями понятия n. Их длины L_m для разных $m>0$, очевидно равны

$$L_m=i_1-1+m(i_2-i_1)+k-i_2+1=$$
$$=k+(m-1)(i_2-i_1),$$

а значит, для произвольной константы L всегда можно подобрать такое $m>0$, чтобы выполнялось $L_m>L$. Полученное противоречие завершает доказательство.

<u>Следствие 1 теоремы 1.</u> Множество именующих нитей не рекурсивного понятия конечно.

<u>Определение 7.</u> Полусхему $S=(N,R,D)$ назовем нерекурсивной, если все ее понятия нерекурсивны.

Ключевым понятием при исследовании полусхем является понятие образца.

<u>Определение 8.</u> Для заданной полусхемы $S=(N,R,D)$ понятие $n \in N$ называется базовым, если $nD=\varnothing$.

Множество таких понятий далее обозначается через N_0 [5].

<u>Определение 9.</u> Для заданной полусхемы $S=(N,R,D)$ именующая нить (n,w) называется терминальной, если $\tau(n,w)\in N_0$ [4, 5].

<u>Определение 10.</u> Для заданной полусхемы $S=(N,R,D)$ и понятия $n \in N$ образцом последнего называется $p=\{(n,w_i)\,|\,i=1,\dots,Q\}$ – конечное множество терминальных именующих нитей n, обладающее следующими свойствами:

1. если $n \in N_0$, то $p = \{(n, e)\}$;
2. если $n \notin N_0$, то существует $f \in nD$ такое, что

   2.1. $\mathrm{im}(f) = h(p)$, $p = \bigcup\limits_{r \in \mathrm{dom}(f)} \{(n, rt(w_i)) \mid r = h(w_i), i = 1, \ldots, Q\}$,

   2.2. для $r \in \mathrm{dom}(f)$ $\{(f(r), t(w_i)) \mid r = h(w_i), i = 1, \ldots, Q\}$ является образцом понятия $f(r)$.

Следствие 2 теоремы 1. Множество образцов не рекурсивной схемы конечно.

Напомним ключевое определение из [4].

Определение 11. Полусхема $S = (N, R, D)$ называется схемой, если для каждого ее понятия существует образец.

Основная теорема о нерекурсивных полусхемах (теорема 2). Всякая нерекурсивная полусхема является схемой.

Доказательство.

Очевидно, что все базовые понятия имеют образцы. Рекурсивно построим последовательность подмножеств множества N по формуле:

$$N_{k+1} = N_k \bigcup \{n \in N \mid (\exists f \in nD) \mathrm{im}(f) \subset N_k\}.$$

Очевидно, что эта последовательность не убывает, а в силу конечности N она стабилизируется, т.е. $N_\infty = \bigcup\limits_{k=1}^{\infty} N_k$ строится за конечное число шагов. Из построения видно, что каждый элемент множества N_∞ имеет хотя бы один образец. Таким образом, полусхема является схемой тогда и только тогда, когда $N = N_\infty$.

Предположим, что это неверно. Тогда найдется $n \in N - N_\infty$ и $f_1 \in nD$ (в противном случае $n \in N_0 \subset N_\infty$).

Рассмотрим $p_1 = \{(n, r) \mid r \in \mathrm{dom}(f_1)\}$ и заметим, что хотя бы для одного $r_1 \in \mathrm{dom}(f_1)$ выполняется $f_1(r_1) \notin N_0$, иначе мы бы получили, что $n \in N_1 \subset N_\infty$. Повторив этот процесс для $f_1(r_1), \ldots, f_1(r_s)$, где $\{f_1(r) \mid r \in \mathrm{dom}(f_1) \wedge f_1(r) \notin N_0\}$, мы сможем подобрать $\{f_{2,1}, \ldots, f_{2,s}\}$ с тем же свойством, что и f_1. Этот процесс должен продолжаться бесконечно, в противном случае мы построим образец понятия n.

Согласно теореме 1, все именующие нити понятия нерекурсивной полусхемы равномерно ограничены по длине. Следовательно, описанный выше процесс не может продолжаться бесконечно, так как он на каждом следующем шаге дает хотя бы одну именующую нить понятия n на единицу длиннее максимума длин именующих нитей, полученных на предыдущем шаге. Полученное противоречие и завершает доказательство теоремы.

Теперь мы можем описать структуру экземпляров, т.е. образцов, каждой именующей нити (n, w) которых поставлено в соответствие значение из домена возможных значений для базового понятия $\tau(n, w)$ (подробнее см. [4]). Таким образом, если рассматривать именующие нити как составные имена атрибутов, то каждый образец понятия дает отношение, представляющее некоторую точку зрения на это понятие. Следовательно, множество примеров объектов предметной области, описываемой не рекурсивной полусхемой, можно трактовать как реляционную базу данных.

Подводя итог, можно отметить следующее. В работе на основании результатов анализа предметных областей различных АИУС в терминах алгебраической метамодели, предназначенной для формальной спецификации структуры предметных областей и предложенной в работах [4 – 5], впервые выделен соответствующий класс полусхем, названный авторами классом нерекурсивных полусхем. Для выделения этого класса проведено разбиение множества понятий полусхемы на подмножества рекурсивных и нерекурсивных понятий. Для понятий сформулирован и доказан критерий их нерекурсивности в терминах алгебраической метамодели. Для представителей выделенного класса нерекурсивных полусхем доказано, что они являются схемами. Последний результат позволяет заменить достаточно трудоемкий алгоритм выяснения является ли полусхема схемой, предложенный и обоснованный в [4], простой проверкой нерекурсивности всех понятий полусхемы.

Полученные в работе результаты позволяют сделать вывод о возможности непосредственного представления нерекурсивных схем посредством реляциионных баз данных. Алгоритм синтеза схемы базы данных по нерекурсивной полусхеме планируется построить в дальнейшем.

Список литературы: ***1.*** Дейт К. Дж. Введение в системы баз данных, 6-е издание. – М.-СПб.-К.: Вильямс, 1999. – 846 с. ***2.*** Грэхем И. Объектно-ориентированные методы. Принципы и практика. Издание третье. – М.-СПб.-К.: Вильямс, 2004. – 879 с. ***3.*** Жолткевич Г.Н., Семенова Т.В. Концептуальное моделирование данных в исследовательских информационных системах средствами реляционных СУБД // Вестник Херсонского гос. техн. ун-та. – 2002, №15. – С. 75-79. ***4.*** Жолткевич Г.Н., Семенова Т.В. К проблеме формализации концептуального моделирования информационных систем // Вісник Харківського нац. ун-ту ім. В. Н. Каразіна. – № 605. – 2003: Математичне моделювання. Інформаційні технології. Автоматизовані системи управління. – Вип. 2. – С. 33-42. ***5.*** Жолткевич Г.Н., Семенова Т.В., Федорченко К. А. Представление полусхем предметных областей информационных систем средствами реляционных баз данных // Вісник Харківського нац. ун-ту ім. В. Н. Каразіна. – № 629, 2004: Математичне моделювання. Інформаційні технології. Автоматизовані системи управління. – Вип. 3. – С. 11-24.

Поступила в редколлегию 03.04.06